EXTREME HEAT EVENT RISK MAP CREATION USING A RULE-BASED CLASSIFICATION APPROACH

Kenneth Rulon Simmons

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Master of Science
in the Department of Geography
Indiana University

December 2011

Accepted by the Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Master of Science.

Daniel P. Johnson, Ph.D., Chair

Aniruddha Banerjee, Ph.D.

Master's Thesis Committee

Jeffrey S. Wilson, Ph.D.

ACKNOWLEDGEMENTS

I wish to thank the members of my thesis committee, Dr. Dan Johnson, Chairman; Dr. Rudy Banerjee; and Dr. Jeff Wilson. As classroom instructors, mentors, and friends, they were the finest, and I very much enjoyed and appreciated the many learning opportunities I had with them. I'm grateful, too, to Joyce Haibe, also of the IUPUI Department of Geography, for her thoughtful and regular input on the thesis process.

I thank Dr. Vijay Lulla and Austin Stanforth, whose regular willingness to share their knowledge of the geospatial sciences and technology greatly aided me with the completion of my thesis work. Several other fellow graduate students also supported me along the way, including Karen Bell, Brian Hanson, Jeremy Webber, Amanda Hamm, Mike Crussel, and others.

Many thanks as well to my father, Rulon Simmons, for both reviewing the initial draft of this thesis and for providing expert advice with the formatting of the final manuscript.

Proper recognition must also be made to the following entities for their financial support of my graduate work: the U.S. Geospatial Intelligence Foundation, the Indiana Space Grant Consortium, the Indiana Department of Workforce Development, and the IUPUI Department of Geography.

Finally, I would most like to acknowledge the love and support of my wife, Roxie, in making the completion of this thesis and advanced degree possible. It was Roxie's encouragement to pursue a new field of study that led me to a very rewarding – and unexpected – new career in geographic

information science. With this thesis now completed, I am looking forward to many more fun times ahead with her and our boys, Jake and Caleb.

ABSTRACT

Kenneth Rulon Simmons

EXTREME HEAT EVENT RISK MAP CREATION USING A

RULE-BASED CLASSIFICATION APPROACH

During a 2011 summer dominated by headlines about an earthquake and a hurricane along the East Coast, extreme heat that silently killed scores of Americans largely went unnoticed by the media and public. However, despite a violent spasm of tornadic activity that claimed over 500 lives during the spring of the same year, heat-related mortality annually ranks as the top cause of death incident to weather. Two major data groups used in researching vulnerability to extreme heat events (EHE) include socioeconomic indicators of risk and factors incident to urban living environments. Socioeconomic determinants such as household income levels, age, race, and others can be analyzed in a geographic information system (GIS) when formatted as vector data, while environmental factors such as land surface temperature are often measured via raster data retrieved from satellite sensors. The current research sought to combine the insights of both types of data in a comprehensive examination of heat susceptibility using knowledge-based classification. The use of knowledge classifiers is a non-parametric approach to research involving the creation of decision trees that seek to classify units of analysis by whether they meet specific rules defining the phenomenon being studied. In this extreme heat vulnerability study, data relevant to the deadly July 1995 heat wave in Chicago's Cook County was incorporated into decision trees for 13 different experimental conditions. Populations vulnerable to heat were identified in five of the 13 conditions, with predominantly low-income African-American communities being particularly at-risk. Implications for the results of this study are given, along with direction for future research in the area of extreme heat event vulnerability.

Daniel P. Johnson, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

BACKGROUND

Significant research has been conducted on the causes and solutions to heat-related morbidity and mortality. Despite the fact that extreme heat causes more mortality in the United States than any other type of weather-related killer (NOAA, 2010), such events rarely receive the kind of public response that other more physically destructive natural disasters do, despite awareness of the heat and its potential danger (Sheridan, 2006). With global warming pointing to an increase in the frequency and intensity of extreme heat events and resultant heat-related morbidity and mortality, the need for comprehensive public health and emergency response is all the greater (IPCC, 2007). Any successful effort towards this end begins with research into both heat wave dynamics and the implications of socioeconomic preconditions of vulnerability.

As a spatially distributed phenomenon, social and environmental factors involved with heat-related health issues are well suited for study via geographic information systems and their related scientific disciplines. Traditionally, studies conducted in this area have sought to identify socioeconomic factors that predispose certain demographic groups to increased risk for heat-related complications. For example, in the late 1990's Morrow posited that, "disaster vulnerability is socially constructed, i.e., it arises out of the social and economic circumstances of everyday living" (Morrow, 1999). The results of this body of research suggest that those with elevated susceptibility for health complications associated with extreme heat include, among other variables, the elderly, those living alone, the urban poor, and those who don't have or use air conditioning

(Semenza et al., 1996). Additionally, individuals with chronic mental disorders or pre-existing medical conditions have an elevated susceptibility to adverse effects of heat waves.

Along with efforts to understand and map the distribution of the physiological, behavioral, and social characteristics associated with extreme heat event mortality, other work has focused on the environmental dynamics of extreme heat events themselves (Dousset et al., 2011). Particular emphasis has been given to understanding the nature and distribution of the urban heat island (UHI) effect, a phenomenon where increased “heat loads” are produced and maintained within areas characterized by a high percentage of heat-trapping man-made materials and a low percentage of natural land cover (Lowry, 1967). Urban centers are typical of such heavily built-up environments, and they can have temperatures of about 1.5°C higher than rural areas surrounding them (Oke, 1995). And even within the urban center itself, the geometry of some neighborhoods or even developments, such as high-rise apartment buildings, can result in small order microenvironments that demonstrate particularly acute tendencies for the UHI effect (Harlan, Brazel, Prashad, Stefanov, & Larsen, 2006).

Recent work in EHE risk assessment has sought to integrate socioeconomic GIS vector data and remotely sensed thermal data for UHI in attempts to discover those areas that are especially vulnerable during heat waves (Johnson, Wilson, & Luber, 2009). Research over the past decade has analyzed such susceptibility at differing scales, incorporating a wide variety of risk factors and methodologies in evaluating their hypotheses. For example, Harlan, Brazel et al. (2006) compared population characteristics, ecological variables, and coping resources in their analysis of extreme heat vulnerability across eight different neighborhoods in the greater Phoenix area. Their findings suggested that heat stress exposure was not even across the neighborhoods, but was greatest in areas inhabited by poor and minority populations. The research also found that these neighborhoods were also burdened with a deficiency of social and material resources to effectively mitigate the heat.

Further studies on heat wave mortality sought to analyze the temperature differences between downtown areas in three Italian cities and their respective airports, where temperature readings are commonly obtained and referenced for an entire city (de'Donato et al., 2008). Results showed a "heterogeneous relationship" between temperatures at the two locations and heat-related mortality, suggesting that the practice of using a single location to assess heat-related risk is inadequate, as the thermal measurement from one location cannot be generalized across an area of disparate land cover types.

Finally, other studies have looked to analyze susceptibility to extreme heat from the perspectives of fields outside of health geography. The work of Whitman, Good et al. (1997) is representative of such efforts, having used traditional regression analyses to determine "excess deaths" beyond a derived baseline during the 1995 Chicago heat wave. The results of the research provided strong evidence of increased vulnerability to heat-related mortality for blacks and the aged, and significantly less risk for Hispanics.

Seeking to build upon the body of work already done in the area of "risk mapping" of heat vulnerability, the present research used knowledge classifiers (also called "knowledge-based classification," "expert systems," "rule-based classifiers," "decision tree classifiers," and "machine-learning approaches") to fuse socioeconomic and environmental data at the Census block group level in an attempt to more precisely identify populations vulnerable to heat-related issues. Better predictions of locations susceptible to heat effects can lead to an improvement of emergency planning before a heat-related crisis develops and the response while it unfolds, conserving scarce resources and saving lives as a result.

Mapping Vulnerable Populations

In identifying and mapping the locations of groups of individuals who might be at higher risk for adverse heat-related complications, two major classes of geographically referenced data have been utilized: vector-based GIS data and remotely sensed raster imagery.

Within geographic analysis, vector formats are commonly used to store large volumes of quantitative information on nearly any type of variable. These data can be both symbolized visually in a geographic information system (GIS) and analyzed statistically. This functionality continues to be effective in mapping the distribution of populations exhibiting heat-related morbidity and mortality risk factors, especially with GIS-ready socioeconomic data being readily available from the U.S. Census Bureau and other sources.

Another type of data highly relevant to epidemiological examinations of EHE issues is the imagery captured by satellite and aerial sensors. These raster images, and particularly the quantitative data behind them, provide opportunities to analyze large geographic areas at a wide range of scales. Specifically in the case of extreme heat events, remotely sensed data offer the ability to estimate land surface temperature (LST) for a multitude of points across an entire region at one instantaneous moment in time. The use of just a single instrument in measuring LST has advantage over that of a network of terrestrial thermometers that can potentially be both inconsistent in their readings and insufficient in their distribution across the region of interest (Dousset, Gourmelon, & Mauri, 2007).

Measurement of land surface temperatures (LST) via remote sensing is possible via thermal band sensors on publicly available imagery (often Band 6 of Landsat data). Such thermal data permit the researcher to quantify and map temperatures across an entire scene, something of particular utility in identifying the urban heat islands that can represent elevated levels of heat-related health risk.

Researching urban climatology and its influence on heat wave risk isn't without its challenges, however. Notably, urban heat islands are complex phenomena, with land surface temperature calculations affected by, among other things, the geometry and multi-dimensional composition of urban surfaces, the angle of the sensor to the surface, and the size of the instantaneous field of view (IFOV) of the sensor (Voogt & Oke, 2003). Additionally, spatial resolution of a particular sensor may not provide the level of detail necessary to properly examine a unit of analysis at a size of interested.

Despite the advantages of both classes of geographic data, discussion in recent years has focused on the need to combine the benefits of both to achieve needed accuracy improvements in classification (Baltsavias & Hahn, 1999). The present research seeks to demonstrate that possibility by showing the development of a heat wave risk map via the fusion of socioeconomic GIS data and remotely sensed environmental indicators.

Knowledge-based Classification

One of the more promising advancements in geographic analysis over the past decade has been the utilization of knowledge classifiers, data mining software that seeks to mimic the thinking and performance of a human subject matter expert in reaching its conclusions (Cohen & Jensen, 1996). Used in such diverse fields as business, medical research, environmental resource management, military intelligence and operations, communications and more (Baijal, Arora, & Ghosh, 2006; King-Sun & Rosenfeld, 1976; Li, Goldgof, & Hall, 1993), knowledge classifiers have the distinct advantage of being able to incorporate dissimilar data types in making assessments. This is done using rule-based logic – essentially *if-then* statements – in narrowing down the list of possible outcomes (Jensen, 2005).

Within digital image processing of remotely sensed data, for example, an analysis would typically start and end with the statistical classification of pixels into clusters of data points with similar brightness values. A knowledge classifier, however, takes that pixel-level classification and adds to it one or more variables – raster, vector, or other data types – and incorporates the additional pertinent information in making a final classification of the pixel. A class membership requirement might dictate, for example, that pixels within the area of interest exceed the threshold of two classification rules – e.g., a brightness value above 200 in the Landsat TM thermal band and elevation below 100 feet – in order to be placed in a particular category. The results of this multi-step, binary rule-based approach is a considerably more complete and robust assessment of the overall area of research than studies focusing on only one specific data type (see Figure 1 for a general representation of the knowledge classifier format).

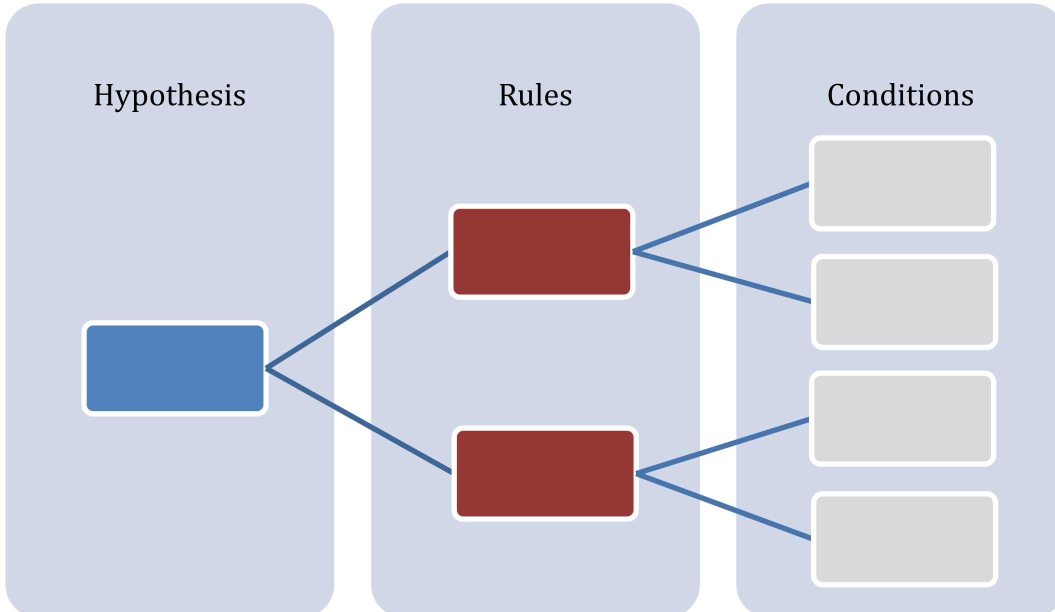

Figure 1. The fundamental building blocks of a knowledge classifier, including hypotheses, rules, and conditions.

Classification via knowledge classifiers has distinct advantages. Foremost among them is the aforementioned ability to apply disparate data types – discrete and continuous – towards a more thorough analysis. As such, knowledge classifiers are a holistic method of data analysis that bridge the gap between geospatial science's two major sub-disciplines, geographic information science and remote sensing, via the integration of vector-based GIS data layers and raster-based imagery.

Rule-based approaches also require no assumption of Gaussian (normal) distribution of the data the way parametric approaches do. Independence of the data is also not assumed.

Finally, knowledge classifiers use a transparent inference process consisting of a chain of binary decisions that can be seen and examined by the user (Jensen, 2005), allowing for a significantly enhanced understanding of the logic behind each classification assignment made by the software.

Knowledge-based classification also has its drawbacks. Among these disadvantages is the creation of overly complex decision trees that results in something called “overfitting,” where the resulting classification is so precise as to only be applicable to the specific conditions (e.g., time and place) of the analysis (Cohen & Jensen, 1996). Such a lack of external validity is typically the result of having incorporated too many variables into the model.

Another potential drawback in using a knowledge-based classification scheme lies in the actual creation of the decision-tree, or knowledge base. Understanding the heuristics underlying one’s classification decisions isn’t as simple as might be expected, even for the expert himself.

Two approaches exist for tackling the creation of such knowledge bases, the first option being for the researcher to independently develop the hypotheses and the supporting rules and conditions that form the hierarchical decision-tree. The second is to enlist the use of automated machine-learning software that uses inductive or deductive inference strategies, such as that incorporated in the commonly used C5.0 algorithm, to determine rules and conditions that satisfy hypotheses (Jensen, 2005).

In order to manually formulate the hypothesis-rule-condition hierarchies that form a decision-tree, a researcher first determines the problem that is to be answered. Based on his or her personal understanding of the research topic – or in consultation with a subject matter expert in the field – the researcher determines the rules that define the hypothesis and the list of conditions that must be met in order for the rule to be true. At times this can be a challenging proposition, however, as subject matter experts themselves can often have a difficult time understanding how it is they come about their conclusions. Enlisting the help of a “knowledge engineer” (i.e., someone adept at eliciting and formatting subject matter expertise for use in a knowledge classifier) can often be helpful in this effort, but the process often takes an excessive amount of time. Given these impediments, the use of automated machine-learning software in determining salient rules and their interaction can be of particular utility.

Despite the potential complexity of manual knowledge base creation, it has become acceptable in recent years for experts to, in fact, develop the decision-tree structure on their own, querying themselves in an effort to successfully parse the rules and conditions for the chosen hypothesis (Jensen, 2005). As outlined in the following methods section, a manual rather than machine-learning approach was taken here in the creation of the knowledge base.

METHODS

The present research examined the greater Chicago metropolitan area for potential susceptibility to heat-related morbidity and mortality. In 1995, Chicago was the center of a catastrophic death toll resulting from a significant extreme heat event (EHE). Nearly 750 individuals perished in a 5-day span, making the heat wave the deadliest environmental disaster in Chicago history. In an effort to better predict and therefore mediate heat wave-related mortality, a model was created for the study incorporating both socioeconomic block group data and environmental data from the time of the 1995 Chicago EHE. This data fusion was accomplished through the use of knowledge-based classification, a holistic approach free of assumptions of Gaussian data distribution commonly made by parametric classification methods.

Data

Data representing values for 25 socioeconomic variables was incorporated into the vulnerability analysis. All data came from the 1990 U.S. Census, the decennial census data information most current to the time of the 1995 heat wave. The data were compiled by fellow IUPUI GISc graduate student Austin Stanforth for use in his thesis involving a principle components analysis of EHE indicators (Stanforth, 2011). As a methodological companion study to Stanforth's work, the present research benefitted from shared data usage and the direct comparison that it allowed.

The 25 socioeconomic variables chosen for this study were ones representative of circumstances traditionally noted in academic literature as ones putting people at risk to adverse effects from excessive heat (Cutter, Boruff, & Shirley, 2003; Morrow, 1999). The factor groups incorporated here include variables associated with race, gender, age, income, living situation, education, and total Census block group population.

The race variables were selected in order to determine the comparative difference in heat risk between the major race groups, including white, black, and Hispanic. Anticipating that some of these differences might be functions of levels of education and financial wealth, variables related to those two factor groups were also inputted into the study. It has been well documented that those living alone, particularly the elderly, have a disproportionately higher rate of mortality during deadly heat waves than do those who live communally (Klinenberg, 2002). Data reflecting both age and living situation were therefore included in the models for this study.

An additional three variables, all environmental – normalized difference vegetation index (NDVI) for measuring the presence of green vegetation; normalized difference built-up index (NDBI) for measuring manmade features; and surface temperature – were added to the 25 socioeconomic variables in the model. These data layers were all derived from 120-meter resolution Landsat 5 TM thermal imagery captured on July 1, 1995, and like the socioeconomic variables, these environmental data sets were compiled by Austin Stanforth for use in his analysis.

An important assumption made in dealing with the data was that all variables were to be considered equal in their potential to predispose a person or place to complications incident to extreme heat. The decision not to weight vulnerability factors is based on the fact that they are typically difficult to rank because they may not exert the same degree of relative influence across space and time (Cutter, et al., 2003; Rygel, O'sullivan, & Yarnal, 2006). Additionally, the primary focus of the present research was the creation of a practical predictive model of comparative heat risk, not necessarily a ranking of the influence of risk indicators, all of which may contribute significantly to heat-related mortality. All 28 socioeconomic and environmental variables in this study were therefore weighted equally.

Preprocessing

The data sets were put through a series of preprocessing steps in order to ready them for use in an expert classifier knowledge base (see Figure 2). First, all socioeconomic variables except median household income and per capita income were standardized by determining the percentage of the Census block group population that was represented in the value of the respective variable. Such standardization was used in several important vulnerability index studies over the past decade and allowed for a true comparison of susceptibility to heat across Census block groups of drastically different population sizes (Cutter, et al., 2003; Voogt & Oke, 2003).

Additionally, several steps were taken to reduce multicollinearity, or the close statistical correlation of two component variables that leads to inaccurate coefficient estimates of individual indicators within the model. A principal components analysis (PCA) was conducted to reduce the large suite of variables down to the factors that can best explain the majority of the variance within the data sets. This mathematical process sought to find the most important information in the data and eliminate data that either contributed little or that was superfluous vis-à-vis other inputs. The result of the PCA was a group of new components that best represented the complete data set. Each new component had associated factor loadings, or correlation coefficients, for each of the original variables against that component (Abdi & Williams, 2010). To improve the interpretability of these components, a varimax rotation procedure was conducted on the PCA results. The varimax algorithm rotated component factor loadings around both axes on a scatterplot, maintaining relative positions of the loadings while simultaneously maximizing the differences between the largest (positive) and smallest (negative) loadings. The result was a simplification of the reading of the PCA results by the way varimax rotation produced a small number of numerically large factor loadings and a large number of small factor loadings for each component outputted in the principle component analysis (Oke, 1995). These large factor loadings were relatively easy to identify, and the variables to which they are linked can be interpreted as those exerting the primary influence over the distribution of values in their respective components. After the principle component analysis results are put through varimax rotation, each of the original variables is typically associated with only one, or perhaps a few, components (Abdi, 2003).

The PCA component values were automatically converted to Z-scores as part of the PCA technique, standardization that was necessary for determining threshold values in the knowledge-based classification process.

The outputs of the principle components analysis were tabular, and as such needed to be converted to the raster format accepted by the classifier software.

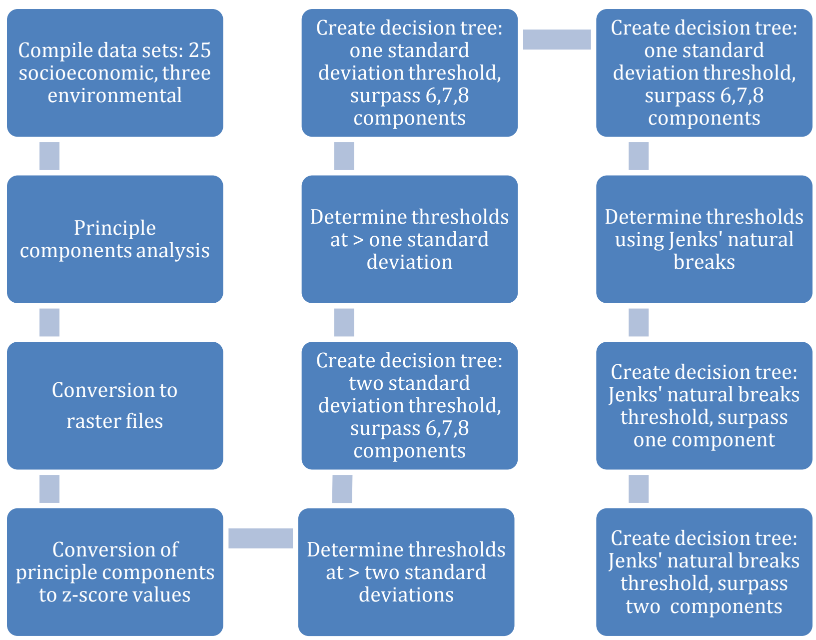

Figure 2. Study methodology, including pre-processing, threshold determination, and decision tree creation.

Knowledge Base Development

Creation of the knowledge bases that were used in the proposed heat wave classification involved the use of the rule-based knowledge engineer and knowledge classifier tools found within a common data processing software package. A fundamental determination made in the process of knowledge base development is the computation of the numerical values that will be used as the variable thresholds upon which a classification will be made. These thresholds (or "splits") are, in fact, the variable-level binary rules against which each pixel is judged in the classification process. For the purposes of this study, standard deviations were used as the primary method for determining these cutoffs, with the value two standard deviations above the mean acting as the high vulnerability threshold for each variable data set. No separate conversion process was needed for the values resulting from the principle components analysis, as this analytic technique calculates deviations from the mean as part of the process, resulting in standard deviation (or "Z-score") values as the outputs.

The use of standard deviation values as an approach to determining thresholds, one detailed in a seminal disease mapping text by Cliff and Haggett (1988), was chosen for several reasons. First, standard threshold values do not necessarily exist within risk management research, primarily because vulnerability factors and their associated threshold levels are not consistent across place and time and therefore don't lend themselves well to a fixed determination. A particular risk factor and its related rule may not possess the same degree of influence in Chicago that it does in Los Angeles, for instance. Second, intervals based on standard deviation take into account the amount of variation there is from the mean, a measure that considers the actual values themselves rather than just the frequency of values that other approaches use. Finally, values lying beyond two standard deviations are by definition outliers among the furthest from the mean, and with regard to the data used in this study, represent conditions most susceptible to heat-related risk.

In order to find the model or models that best portray heat susceptibility, several other conditions were tested that varied both the threshold level and the number of components within which a Census block group had to surpass the threshold in order to be classified vulnerable. Additionally, an alternative method for threshold determination was used in the form of the natural breaks methodology pioneered by American cartographer George Jenks. Sometimes called the Jenks Optimization Method, this statistical technique seeks to determine the best intervals within a range of values by finding the groups of values that minimize the deviation from the mean within the classes while maximizing the variance between them. In this way, intervals are dictated by the natural distribution of the values themselves rather than a more arbitrary method. This approach to creating class intervals has become widely accepted, so much so that the most popular software in the field of geographic information science uses it as its default method for such processes. Given this, several conditions based on the natural breaks methodology were included in the study to serve as a useful comparison to other experimental conditions in the search for the best representation of heat wave vulnerability in the Chicago research area.

RESULTS

In preparation for the knowledge-based classification, the standardized data were subjected to a principle components analysis (PCA) in a common statistical software package. The PCA with varimax rotation produced 27 total components, a figure nearly equal to the number of variables incorporated into the analysis. A review of the components' eigenvalues allowed for a determination of which these components to retain and which to eliminate. The most common rule of thumb in this regard is to keep components with eigenvalues over 1.00 and discard all others as "noise" (Kaiser, 1960). Using this criterion, eight components were kept for incorporation into the knowledge classifier decision tree. Collectively these top eight components accounted for 69.17% of the variance in the variables.

Review of these eight components revealed that the second and fourth components were negatively correlated with vulnerability to extreme heat and would therefore require thresholds set below rather than above the mean to determine heat risk vis-à-vis that component (see Table 1).

Thresholds of Two Standard Deviations; Six, Seven, and Eight Component Requirements

Vulnerability at the Census block group level was assessed starting with the classification parameters sensitive to the highest risk. A decision tree was created in a knowledge engineer tool accordingly, setting threshold levels at two standard deviations above the mean. In so doing, a Census block group would be classified as vulnerable for a component only if it had a value greater than two standard deviations above the mean, or below for components negatively correlated with susceptibility to heat. With thresholds, or "splits", set at two standard deviations, the first decision tree made required Census block groups to exceed this two standard deviation cutoff for all eight components. Any block group meeting these parameters would be at higher risk for adverse heat-related effects. No Census block groups met the requirements of this condition, however. Classifications were next made stipulating block groups exceed thresholds of any seven components at the two standard deviation cutoff level, but here, too, no Census block groups met the requirements. No vulnerable areas were identified under the six component condition as well.

Thresholds of One Standard Deviation; Six, Seven, and Eight Component Requirements

With no block groups surpassing the thresholds of six, seven, or eight components when set at two standard deviations, splits were next set at one standard deviation from the mean. The Cook County area of interest was again classified by looking for Census block groups that surpassed this new cutoff value for all eight components. As with the test run with a two standard deviation split, no Census block groups surpassed one standard deviation thresholds for all eight components, nor did they for six or seven component sets at one standard deviation.

Thresholds of Zero Standard Deviation; Six, Seven, and Eight Component Requirements

Continuing the analysis, thresholds were set at the mean, or a standard deviation of zero, effectively splitting the component values into halves of positive and negative Z-values. A decision tree was again created classifying each Census block group as vulnerable if it exceeded this new threshold value for each of the eight components. In this instance, 31 Census block groups met this requirement. The same block groups all met the classification requirements of exceeding six and seven component thresholds as well, with no difference between the conditions in terms of numbers and locations of Census block groups classified as at-risk [Figures 4, 5, and 6].

Thresholds of Two Standard Deviations, One and Two Component Requirements

With 31 Census block groups identified as being locations of populations at-risk for heat-related complications, additional analytic conditions were tested to find a more general prevalence of vulnerability. To this end, two standard deviation thresholds were reinstituted into the knowledge classification model, but with a requirement of surpassing only one of the eight components. The result was 546 Census block groups exceeding this threshold parameter for any one of the eight components created in the PCA analysis.

Next, the area of interest was analyzed for block groups surpassing two standard deviation thresholds for any two components. The outcome was that nothing met this requirement and no vulnerability was identified under these parameters.

Thresholds based on Jenks' Natural Breaks, One and Two Component Requirements

The final two conditions tested involved the use of the natural breaks methodology for determining threshold values developed by George Jenks. The values for each of the eight components were grouped into four interval classes via this method, and the value defining the near end of the class furthest from the mean was used as the quantitative cutoff in the model. A decision tree was created stipulating Census block groups exceed one of these new thresholds for any one component. The result was a risk map where 2096 block groups were classified as vulnerable based on these criteria. No Census block groups were classified as vulnerable based on a second decision tree requiring that they exceed natural breaks thresholds of any two components.

DISCUSSION

This study sought to use knowledge-based classification as a new and practical approach for analyzing socioeconomic and environmental data related to human susceptibility to extreme heat events. The final product of this effort was to be a "risk map" classifying vulnerable populations by Census block groups in a research area defined by Chicago's Cook County. When all 13 of the experimental conditions were run, risk maps of varying levels of specificity resulted, leading to several distinctly different pictures of comparative vulnerability in the area researched.

The first condition run in the study was the most constraining as it sought to pinpoint Census block groups that fell beyond the second standard deviation in the value distributions of all eight components representing grouped vulnerability factors. Any Census block group that met these requirements would be considered at extreme risk in the event of a heat wave and indeed would have been at the highest level of comparative risk of any Census block groups classified vulnerable among all conditions tested in this study. No Census block groups met these demands, however, resulting in a blank risk map. That no vulnerability was identified in this condition, nor in the six and seven component experimental conditions at two standard deviations, is not surprising in hindsight, given the model's sizeable number of component and extreme threshold requirements. The *a priori* assumptions of this study were that at least a few Census block groups would meet the requirements of these first three experimental conditions. But as at-risk as the population of a particular Census block group may be, the results of this study lead one to conclude that no group of people will ever be considered vulnerable based on all or even many factors of susceptibility. Indeed, this was also the case with the six, seven, and eight component conditions at the less discerning one standard deviation threshold level.

An examination of the distributions for the values of each of the eight PCA components provides some insight into why no Census block groups were classified as vulnerable under these first and most constraining six experimental conditions. In looking at their histograms, one sees that, 1) the mode within the distributions sometimes account for up to about 4,000 of the 4,174 total Census block group values, or 96%; 2) the distributions are sometimes rather skewed, and; 3) the tails are long and flat on either side, accounting for a very small number of the total values (see Figures 4 and 5). With values for as few as seven block groups falling beyond the second standard deviation of some components, the probability of a block group exceeding thresholds set at this level for all eight components calculates to 1.025200974798e-16, or essentially zero. At least one other study has sought to map risk using principle components analysis and failed to find any geographic unit that fell into the highest categories of vulnerability for all components (Voogt & Oke, 2003).

Having tested six, seven, and eight component conditions at both two and one standard deviation thresholds and failed to identify any vulnerability, the study moved to testing conditions based on thresholds set at a zero standard deviation from the mean. Thresholds based on the median involved splitting the range of values for each component equally in half, meaning half of the values for the component fell above this threshold and half fell below. When measured against this term, 31 Census block groups fell into the halves of the distributions representing vulnerability for all eight components. These same Census block groups and no others also exceeded the zero standard deviation thresholds for six and seven components.

As a method of determining risk, zero standard deviation thresholds are clearly not as selective as those set at one and two standard deviations. By definition, thresholds not deviating from the center point of a distribution don't represent the values found at the tails of a histogram that are typically associated with increased hazard risk. Still, with no preexisting knowledge of how population demographics would compare against the eight components that resulted from the principle components analysis, finding vulnerability was necessarily a process of elimination, one that in this case required nontraditional parameters to identify relative susceptibility. That 31 Census block groups fell into the vulnerable halves of distributions for as many as eight composite factors of vulnerability is significant, and though few in number, the geographic areas classified in this condition represent places of uniquely elevated potential risk to heat waves.

A more general trend of susceptibility to heat began to emerge when the requirement for the number of thresholds exceeded was reduced from the maximum, eight components, down to the minimum, one component. It could be assumed that mandating that Census block groups only exceed a two standard deviation threshold in but one of the eight components should result in more than the 31 block groups identified in previous iterations, and in fact it did. Approximately 546 Census block groups met this set of conditions, painting a picture of more widespread susceptibility than previous conditions. Likewise, the one component model for Jenks' natural breaks-determined thresholds also resulted in a much broader classification of vulnerability across the area of interest, one that classified considerably more Census block groups vulnerable than the model with one component at two standard deviations.

A better understanding of the dynamics of vulnerability at play in this study can be gained via a closer examination of the results of the experimental condition that had thresholds determined by Jenk's natural breaks. First, when one looks at the risk map (see Figure 10) that resulted from this experimental condition, it is clear that vulnerability by Census block groups was regionalized rather than random. The areas classified vulnerable in this condition were largely clustered in very homogenous groupings, ones that showed a distinct divide between the urban areas in and around downtown Chicago and the less densely built-up areas of suburban Cook County. Even more telling was the side-by-side comparison of this risk map with the distribution of Component One from the principle components analysis (see Figure 3). The area identified as vulnerable using the Jenks' natural breaks-determined thresholds closely mirrored the regions within Component One showing the greatest concentration of the variables placed within it. Those variables included ones related to predominately African-American communities characterized by a high percentage of families living below the poverty line, a finding consistent with those of other research conducted on this subject (Harlan, Brazel et al. 2006; Whitman, Good et al. 1997). This finding strongly suggests that although extreme heat may disproportionately affect the already disadvantaged African-American community, it can also be better addressed as a public health issue because a significant portion of the overall at-risk population is well-defined.

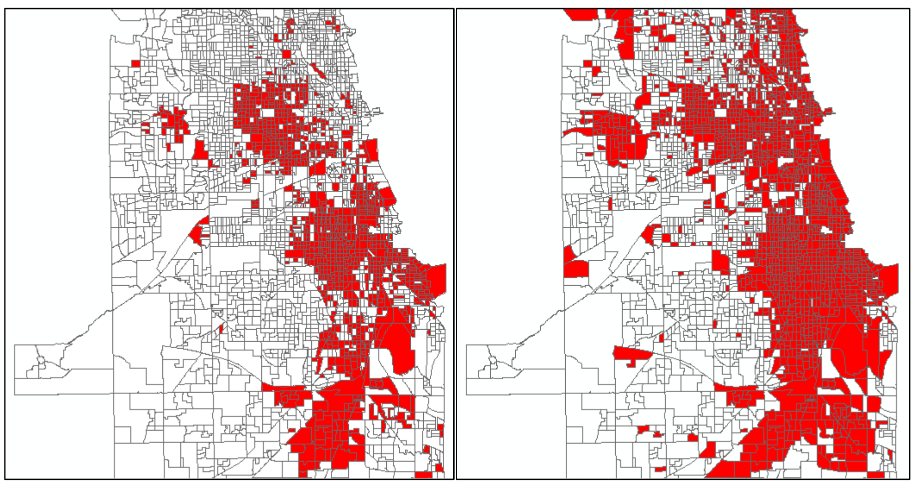

Figure 3. Comparison of Component One, left, and the risk map for Census block groups exceeding thresholds determined by Jenks' natural breaks, right.

The results of this study have some additional practical implications for public policy. Understanding the geographic extent of extreme heat susceptibility naturally improves the budgeting, planning, and execution of emergency management contingencies. For example, with detailed information on at-risk blocks or neighborhoods, emergency planners can make better decisions on locating mobile cooling centers or extending hours for makeshift heat refuges such as public libraries, park facilities, or police stations.

Improvements can also be made to ongoing mitigation and community awareness efforts, including those coordinated with private sector partners.

Accurate heat wave risk maps also have inherent advantages for the response to extreme heat events when they do occur. Both time and resources – equipment, manpower, and monetary – can be focused to where they are most urgently needed. Visits to the isolated elderly can be targeted, as can be warning telephone messages, text messages, and e-mails. In terms of the results of this study, the 31 Census block groups identified in the conditions using zero standard deviation-thresholds would be particularly good places to emphasize such efforts, given their above average placement in all eight vulnerability components.

CONCLUSION

This study showed that as a method for approaching large scale disaster planning and research, knowledge-based classification works well. Both vector and raster data formats were seamlessly incorporated into the model as socioeconomic and environmental variables, respectively. What resulted from the models were practical indications of geographic risk potential to heat wave phenomena. While the specific results of this experiment only have validity to Chicago in 1995, the methodology nevertheless has valid application beyond the parameters of the present study.

Perhaps most intriguing among the results of the various experimental models was the complete absence of any Census block groups being classified as vulnerable under any conditions requiring said block groups to surpass the thresholds of any two components. It would seem, however, that the nature of the aforementioned principle components histograms made it virtually impossible for any Census block group to surpass the threshold requirements of all eight components, despite it seeming reasonable that at least a few block groups would align with two or more components. That vulnerability was broadly identified in the one-component models but not at all in the two-component models is notable, with the reason likely lying in the nature of the way principle components are determined.

Principle components analysis (PCA) seeks to take many possibly correlated variables and regroup them into new components that are not strongly correlated with each other. Thus, finding Census block groups that align well with multiple, orthogonal components isn't likely, at least not in the regions beyond one or two standard deviations of variance. A block group that met stringent threshold requirements for one component simply did not – or perhaps could not – meet the same requirements for any other component. The 31 Census block groups that met the requirements of six, seven, and eight components in an earlier stage of the study weren't necessarily an exception to this assertion, but at .7% of the total count, these block groups were the relative few that met the criteria of fairly liberal thresholds set at a standard deviation of zero. The unanticipated result of having no block group exceed thresholds for more than one component speaks to the complexity of heat wave vulnerability and the methods for examining it.

Future research in this area might consider some modifications to the methodology of the present study. The 28 vulnerability variables in this experiment were a considerable amount, but inclusion of others in future studies might offer some increased explanation over the current model. For example, in the midst of an economic recessionary era such as that currently experienced in the United States, unemployment figures might represent a shorter term risk with heat waves that is real but less permanent than that represented by indicators tied to poverty rates. Conversely, some of the Census variables incorporated in the present model might have been superfluous, offering little added explanation of heat risk in the area of analysis beyond that of key variables. The splitting of certain variables into separate figures for males and females might have been such a case within this study. For example, the male and female data sets for children aged five years and younger were both placed in Component One by the algorithm used in the principle components analysis software, suggesting that they were very correlated to each other and that by being aggregated instead of separated, they might have improved the parsimony of the model.

The current experimental model did not use a data smoothing technique such as the kernel density function, something that future studies along these lines will likely benefit from. The use of kernel density functions is common in addressing the artificial nature of organizing data by census or political boundaries. The process identifies through the use of contours the density of point and line data. It does so by taking values of and distances to adjacent features into account in order to estimate the pixel-level values between and around them. The resulting surface is a smooth gradient between adjacent features that cuts across demographic (e.g., Census) and political boundaries, thus providing a visualization of the dispersion of the phenomenon that is independent of man-made methods of data aggregation (see Figure 11 for an example of kernel density estimation).

Subsequent research should also consider the use of spatial statistics, mathematical algorithms that account for spatial autocorrelation, or co-variance in the values of geographic areas topologically connected. Much like the kernel density function, spatial statistics methods take into account the data from surrounding units of analysis in order to get a more representative figure for each one of them. This methodology is appropriate for the type of analysis conducted in the present heat wave vulnerability research, but as a complex procedure requiring significant effort in its own right, spatial statistics lied outside the scope of this study and was not included in the model.

Finally, future studies in this area might consider methodologies that alternatively provide the possibility for a more multidimensional perspective on vulnerability to heat waves. In this study, incorporating principle components analysis results into a knowledge classifier yielded risk maps that showed vulnerability based on Census block groups that almost uniformly exceeded risk thresholds for but one component. While block groups each might be color-coded by the PCA component for the one threshold it exceeded – thus providing further insight into factors driving vulnerability at that block group – other methodological approaches might allow for assessments of cumulative vulnerability that would permit ranking of susceptibility to heat waves by the number of known risk factors a Census block group exhibits. Using original variable values rather than principle components, albeit fewer than the 28 total included in this study, could potentially result in Census block groups demonstrating cumulative vulnerability of differing levels in a knowledge-based classification. Additionally, traditional statistical methods such as multivariate regression analysis could offer informative insight into this research area, providing factor weights that quantify each variable's impact on heat wave vulnerability, as well as a regression equation that allows for prediction of each Census block group's overall risk based on its values for each variable. An approach such as this could provide a powerful method of determining which populations will be affected during an extreme heat crisis and by how much.

Continued efforts into researching extreme heat vulnerability – the nation's leading weather-related mortality risk – can lead to greater understanding of the problem, more efficient and effective emergency planning, and ultimately reduced loss of life.

APPENDIX

Table 1. Variable names and descriptions.

File Name	Description
TOTAL_POP	Total population: Total
WHITE_POP	Total population: White alone
BLACK_POP	Total population: Black or African American alone
AIAN_POP	Total population: American Indian and Alaska Native alone
ASIAN_POP	Total population: Asian alone
NHPI_POP	Total population: Native Hawaiian and Other Pacific Islander alone
OTHER_RACE	Total population: Some other race alone
HISPANIC_P	Total population: Hispanic or Latino
MALE_5UNDE	Total population: Male; 5 years and younger
MALE_65UP	Total population: Male; 65 years and older
FEMALE_5UN	Total population: Female; 5 years and younger
FEMALE_65U	Total population: Female; 65 years and older
M65UP_LIVE	Over 65 years: In households; In nonfamily households; Male householder; Living alone
F65UP_LIVE	Over 65 years: In households; In nonfamily households; Female householder; Living alone
UP65_GRPLI	Over 65 years: In group quarters; communal living
MALE_NHSD	Over 25 years: Male; Educational attainment; below high school (No High School Degree)
MALE_HSD	Over 25 years: Male; High school graduate; includes equivalency (High School Degree)
FEMALE_NHSD	Over 25 years: Female; Educational attainment; below high school (No High School Degree)
FEMALE_HSD	Over 25 years: Female; High school graduate; includes equivalency (High School Degree)
MHI_1999	Households: Median household income in 1999
MFI_1999	Families: Median family income in 1999
PCI_1999	Total population: Per capita income in 1999
BELOWPOVER	Population with poverty status determined: 1999 Income below poverty level
POV_5UNDER	Population with poverty status determined: 1999 income below poverty level; 5 years and under
POV_65UP	Population with poverty status determined: 1999 income below poverty level; 65 years and over
TEMP_MEAN	Approximate surface temperature of residential spaces from Landsat 5 Thermal band
NDBI	Normalized Difference Built-up Index estimate of built environment within residential space
NDVI	Normalized Difference Vegetation Index estimate of built environment within residential space
MORTALITY	Quantity of heat-related mortalities within the specified boundary shapefile

Table 2. Factor loadings for eight retained varimax-rotated factors with most significant positive (yellow) and negative (pink) correlation to heat wave vulnerability.

Rotated Component Matrix[a]

	Component							
	1	2	3	4	5	6	7	8
pct_blwpov	.849	-.006	.044	-.219	-.116	-.014	.094	-.023
pct_pov5un	.833	.133	-.022	-.006	-.116	-.052	-.092	.101
pct_white_	-.776	.221	.153	.170	.018	-.011	-.287	.097
pct_black_	.713	-.526	-.077	.004	.003	-.018	.263	-.157
mfi_99copy	-.688	-.156	.011	.311	-.462	-.059	-.158	.080
mhi_99copy	-.672	-.118	-.113	.396	-.410	-.075	-.150	.067
pci_99copy	-.632	-.217	.072	.137	-.538	-.027	-.165	.007
pct_fem5un	.525	.352	-.122	.267	-.068	-.087	-.124	.204
pct_male_5	.498	.354	-.167	.230	-.081	-.080	-.115	.053
pct_hispan	.053	.919	-.115	-.225	.014	-.008	.063	-.019
pct_other_	.078	.890	-.108	-.208	-.001	-.006	.083	-.048
pct_f65alo	-.088	-.070	.857	-.044	.145	.048	-.114	.013
pct_fem65u	-.190	-.188	.826	.089	.259	.088	-.017	-.008
pct_pov65o	.324	-.113	.569	-.209	-.131	.192	.187	-.026
pctgrplv65	-.026	.076	.315	.037	-.204	-.043	.182	.002
ndbi_copy	.188	.164	.019	-.871	.003	-.002	.035	.109
ndvi_copy	-.170	-.255	.012	.871	.009	-.006	-.055	.044
pct_m_hsd	-.062	.010	-.032	.002	.763	.031	.156	.035
pct_f_hsd	-.060	-.102	.366	.123	.743	-.039	-.099	.035
pct_m65alo	.024	-.049	.212	-.053	-.028	.895	.042	.001
pct_aian_p	.029	.106	-.176	-.021	.000	.849	-.077	.028
pct_male65	-.211	-.181	.419	.119	.133	.631	.084	-.033
pct_nhpi_p	-.113	.123	.011	.098	-.128	-.051	.576	.156
pct_f_nhsd	.305	-.083	.105	-.064	.158	.061	.560	-.006
pct_m_nhsd	.185	-.009	-.022	-.159	.228	.021	.505	-.067
temp_copy	.076	.029	.012	-.008	.082	.026	.148	.861
pct_asian_	-.249	-.161	-.038	-.354	-.093	-.026	-.201	.402

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 14 iterations.

Table 3. Eight PCA components with their associated original variables, those both positively correlated with heat wave vulnerability and those negatively correlated (*italics*).

Component One	Component Two	Component Three	Component Four
Percent Below Poverty Level	*Percent of Population Hispanic*	Percent of Population Female, 65+, Living Alone	Normalized Difference Vegetation Index (NDVI)
Percent 5 and Under, Below Poverty Level	*Percent of Population Other*	Percent of Population Female, 65+	*Normalized Difference Built-up Index (NDBI)*
Percent of Population Black		Percent of Population 65 and Up, Below Poverty Level	
Percent of Population Female 5 and Under		Percent of Population, 65+, Communal Living	
Percent of Population Male 5 and Under			
Percent of Population White			
Median Family Income			
Median Household Income			
Per Capita Income			
Component Five	**Component Six**	**Component Seven**	**Component Eight**
Percent Male with High School Diploma	Percent of Population Male, 65+, Living Alone	Percent of Population Native Hawaiian, Pacific Islander	Temperature, Approximate
Percent Female with High School Diploma	Percent of Population Am. Indian, Alaska Native	Percent Female with No High School Diploma	Percent of Population Asian
	Percent of Population Male, 65+	Percent Male with No High School Diploma	

Table 4. Results of 13 test conditions.

Number of Components to Exceed	Threshold	Result
Six	Two Standard Deviations	No Census block groups identified as vulnerable
Seven	Two Standard Deviations	No Census block groups identified as vulnerable
Eight	Two Standard Deviations	No Census block groups identified as vulnerable
Six	One Standard Deviation	No Census block groups identified as vulnerable
Seven	One Standard Deviation	No Census block groups identified as vulnerable
Eight	One Standard Deviation	No Census block groups identified as vulnerable
Six	Zero Standard Deviations	31 Census block groups identified as vulnerable
Seven	Zero Standard Deviations	31 Census block groups identified as vulnerable
Eight	Zero Standard Deviations	31 Census block groups identified as vulnerable
One	Two Standard Deviations	546 Census block groups identified as vulnerable
Two	Two Standard Deviations	No Census block groups identified as vulnerable
One	Jenks' Natural Breaks	2096 Census block groups identified as vulnerable
Two	Jenks' Natural Breaks	No Census block groups identified as vulnerable

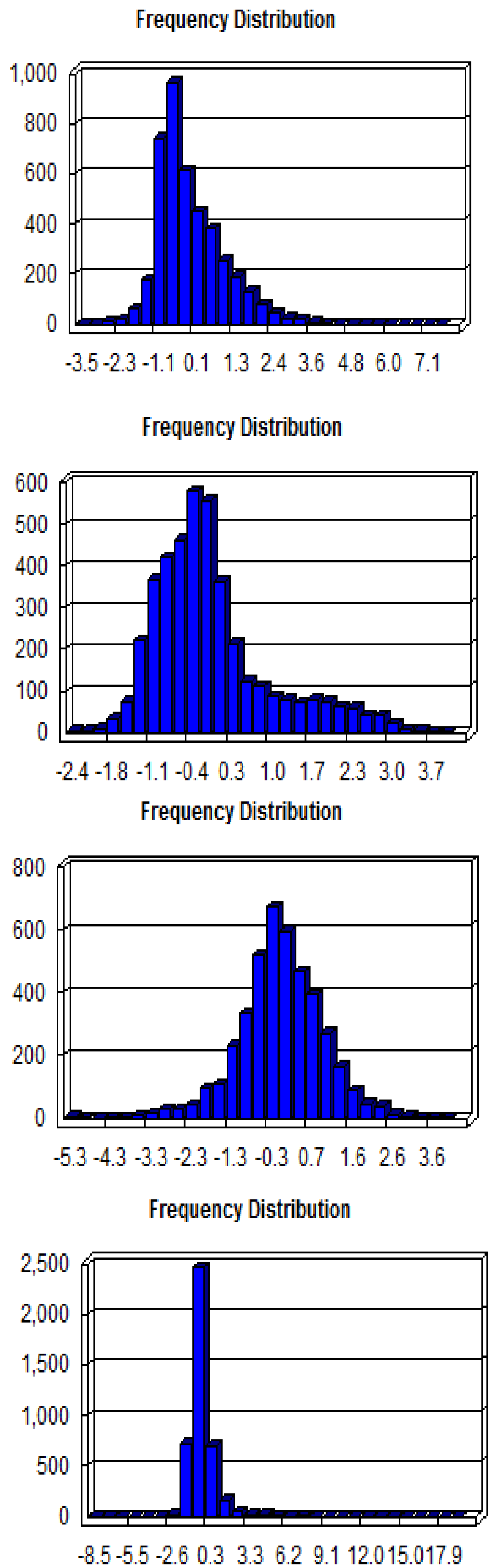

Figure 4. Frequency distributions, Components One through Four, in descending order.

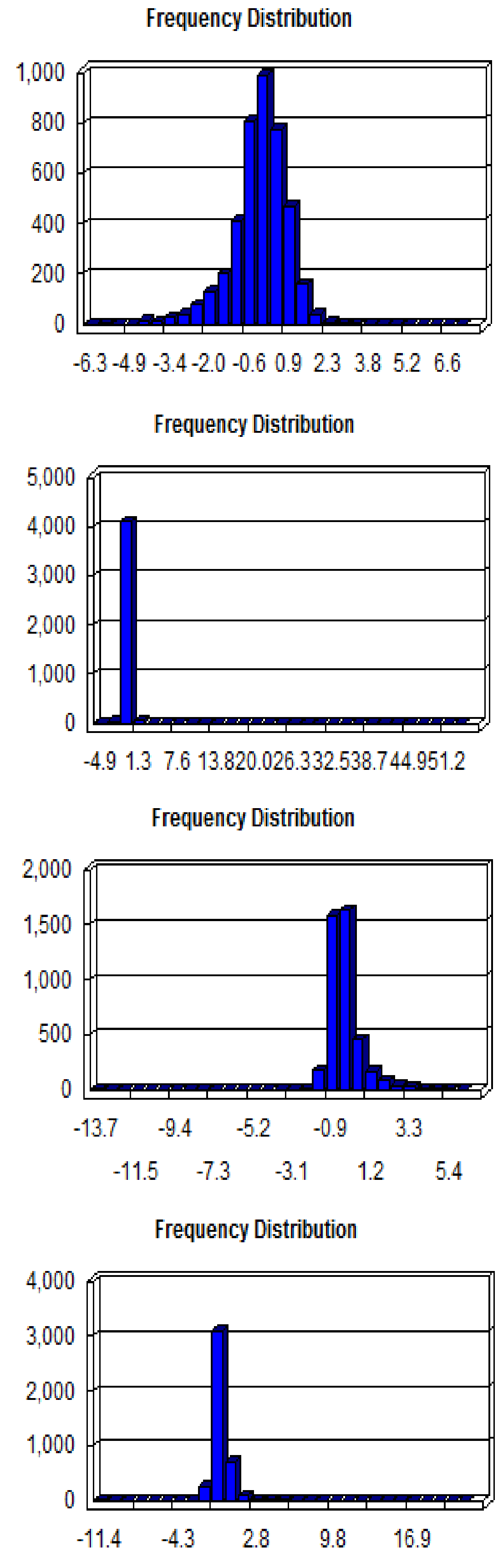

Figure 5. Frequency distributions, Components Five through Eight, in descending order.

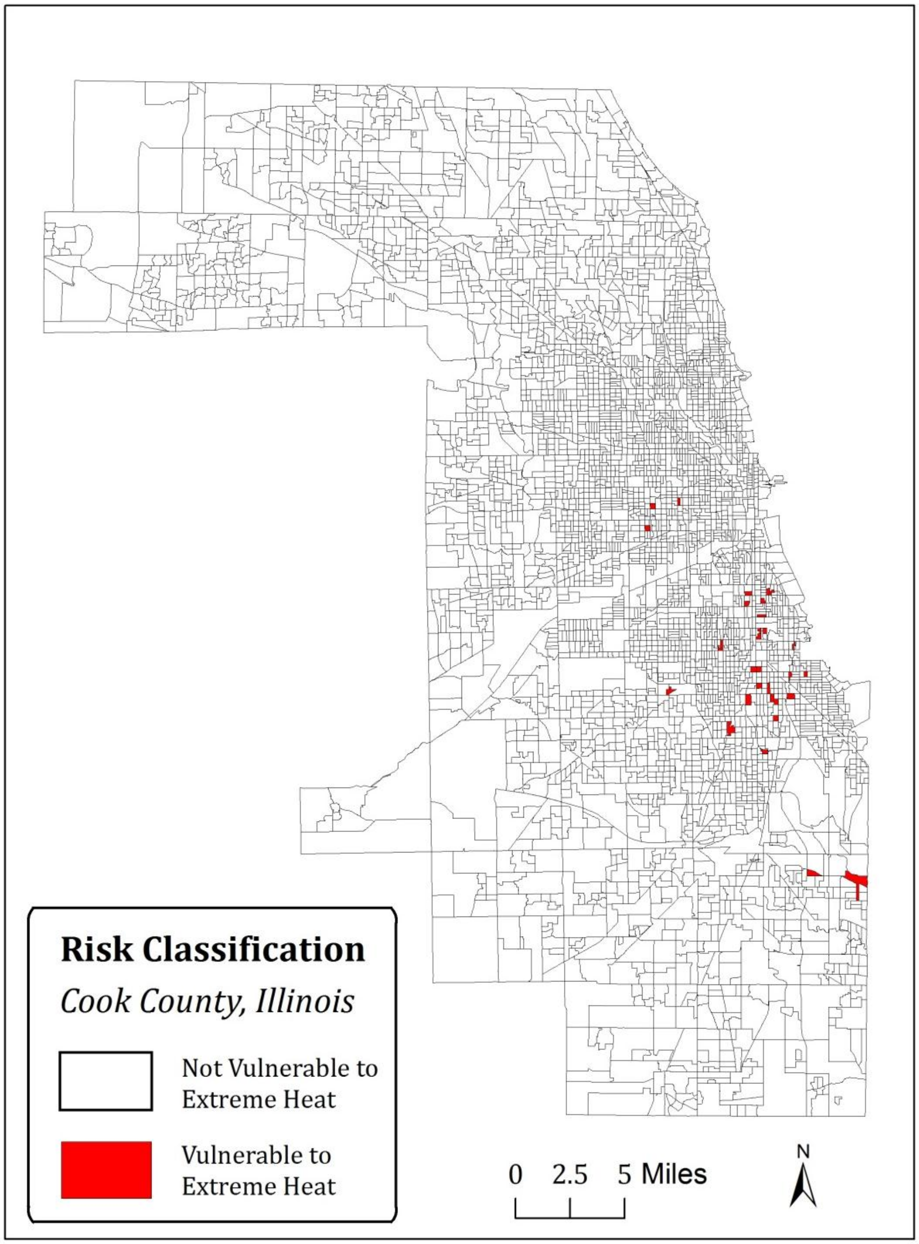

Figure 6. Risk map – eight components, thresholds set at zero standard deviation.

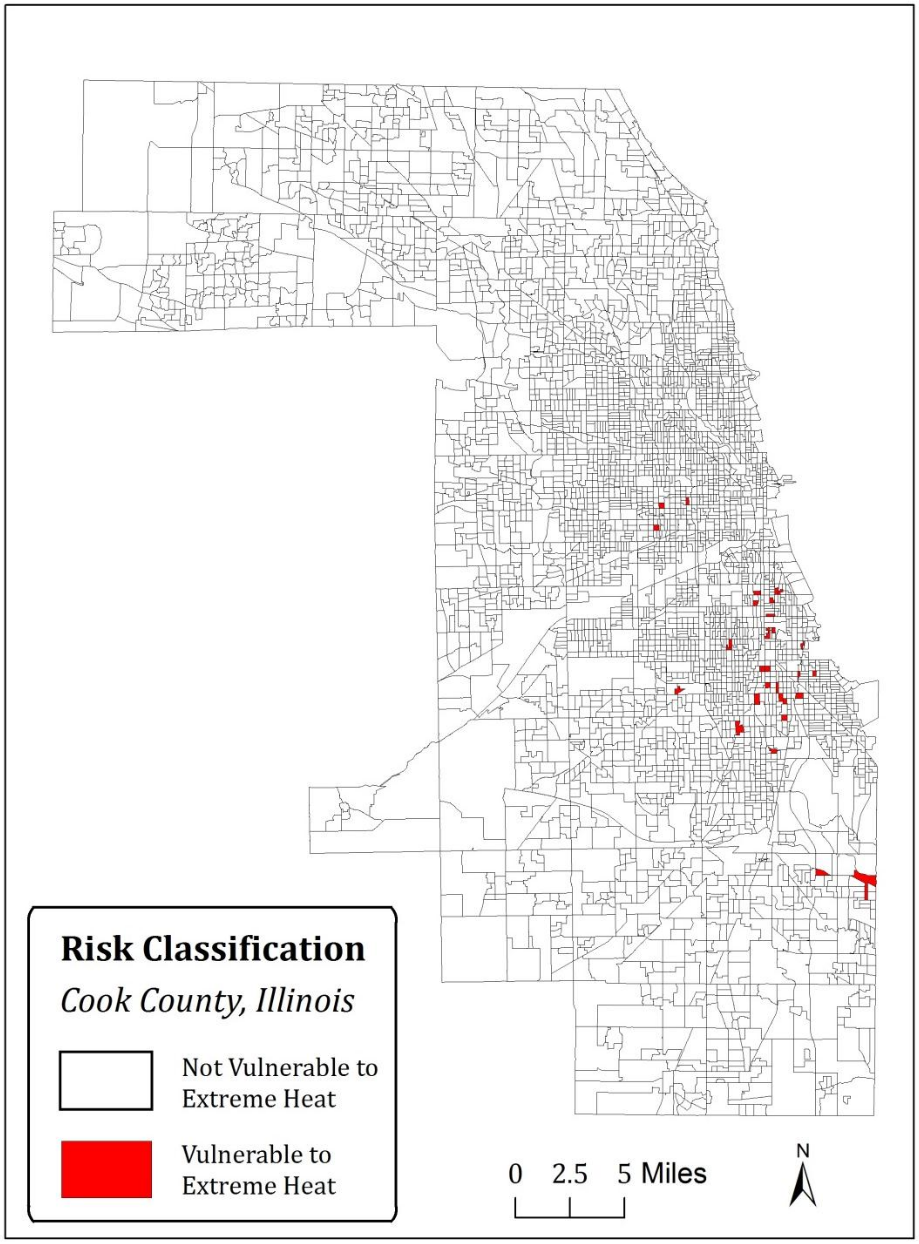

Figure 7. Risk map – seven components, thresholds set at zero standard deviation.

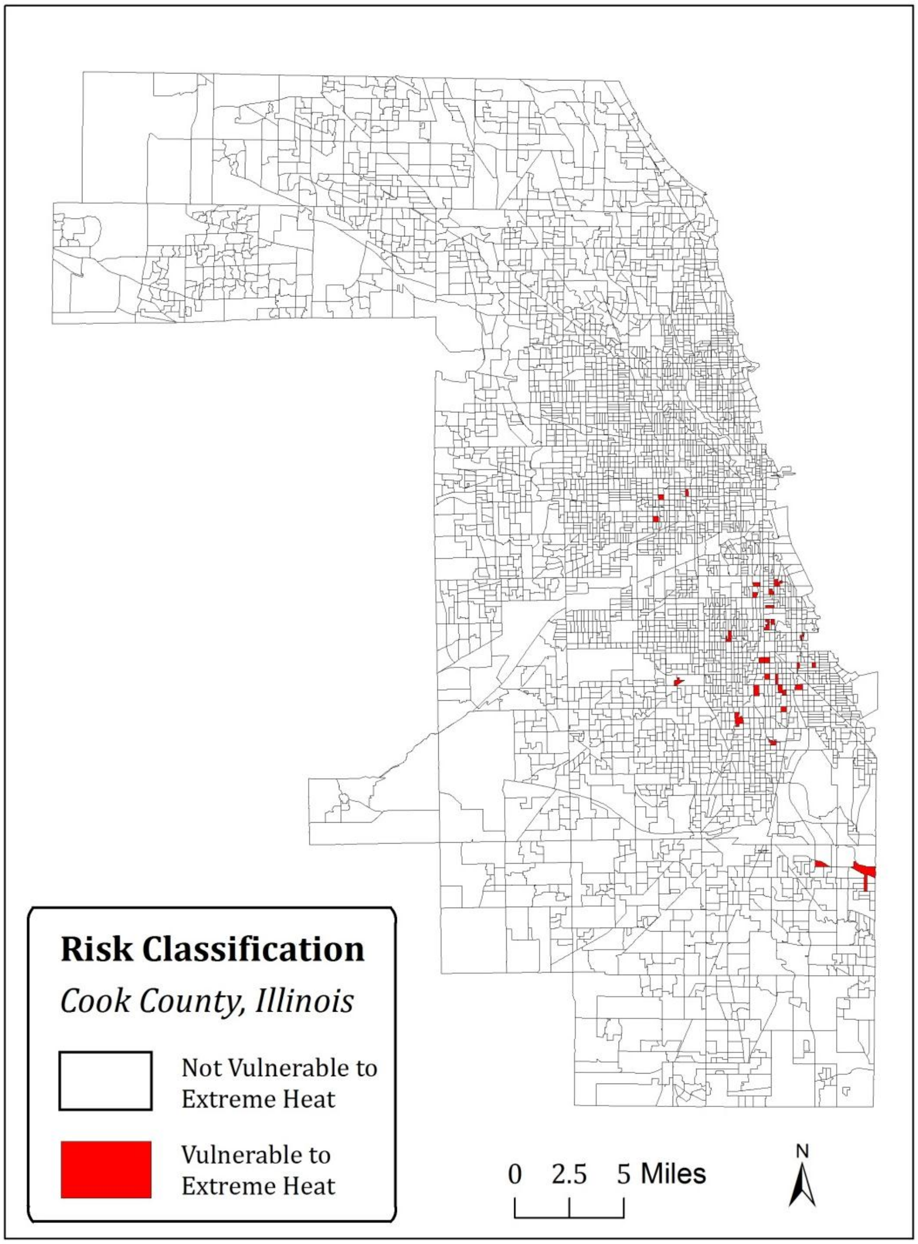

Figure 8. Risk map – six components, thresholds set at zero standard deviation.

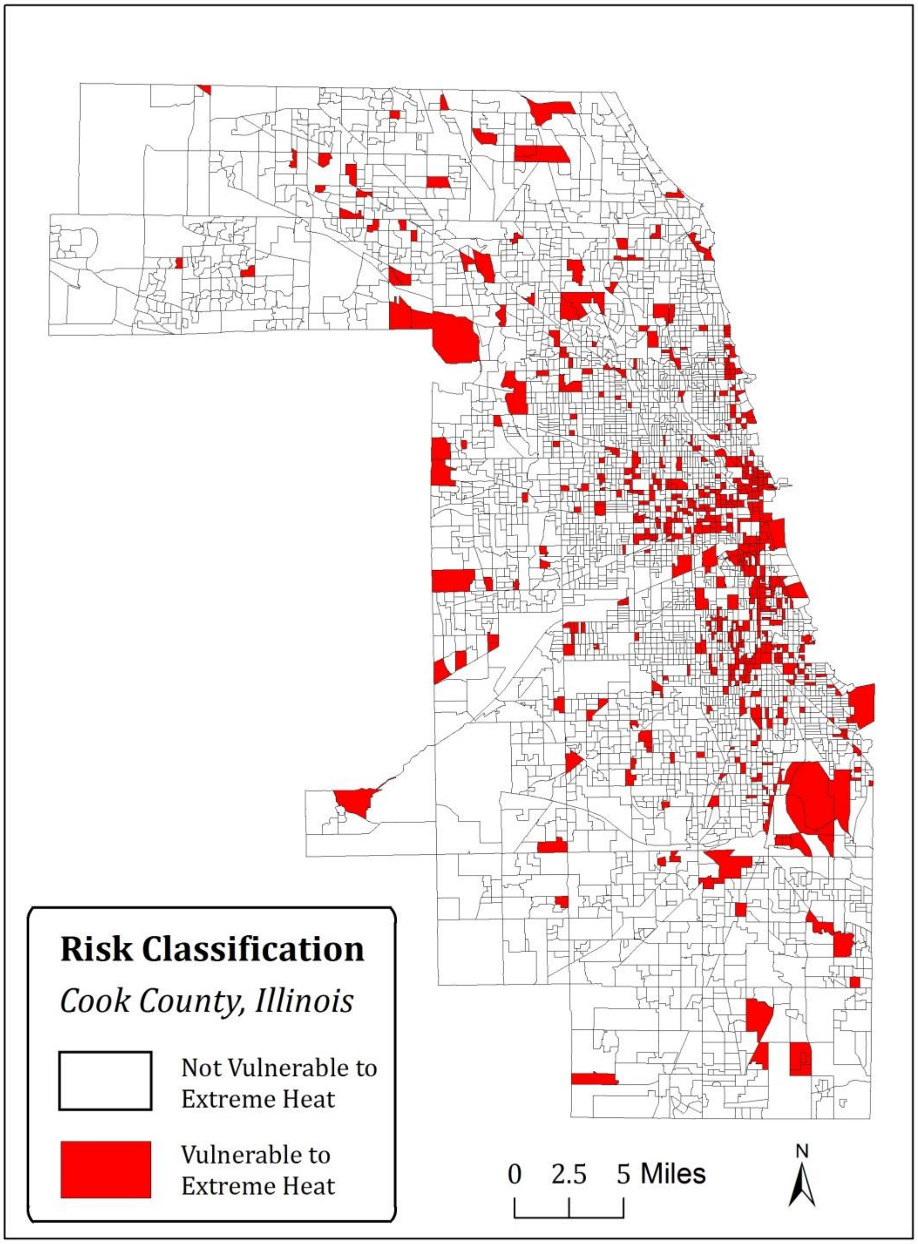

Figure 9. Risk map – one component, thresholds set at two standard deviations.

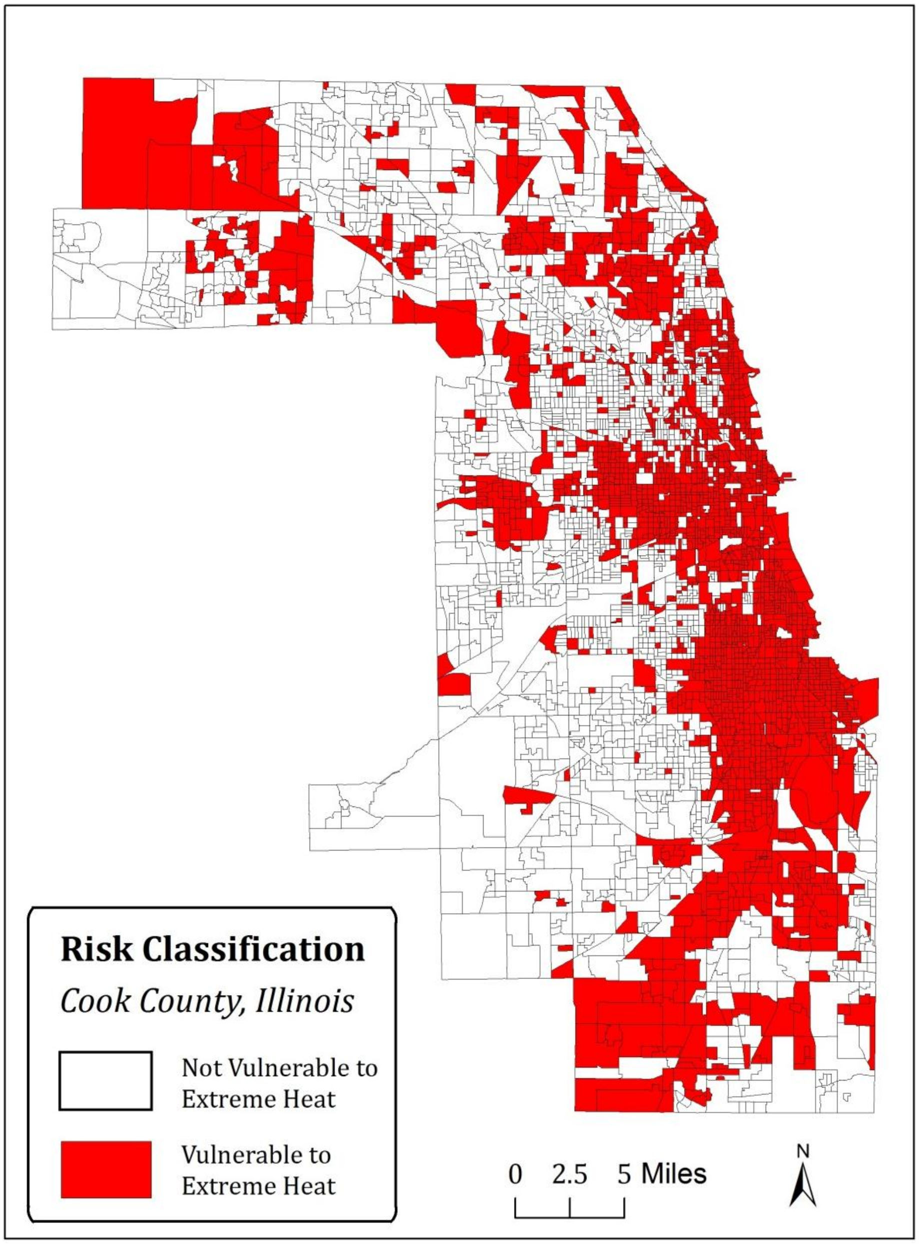

Figure 10. Risk map – one component, thresholds set via Jenks' natural breaks.

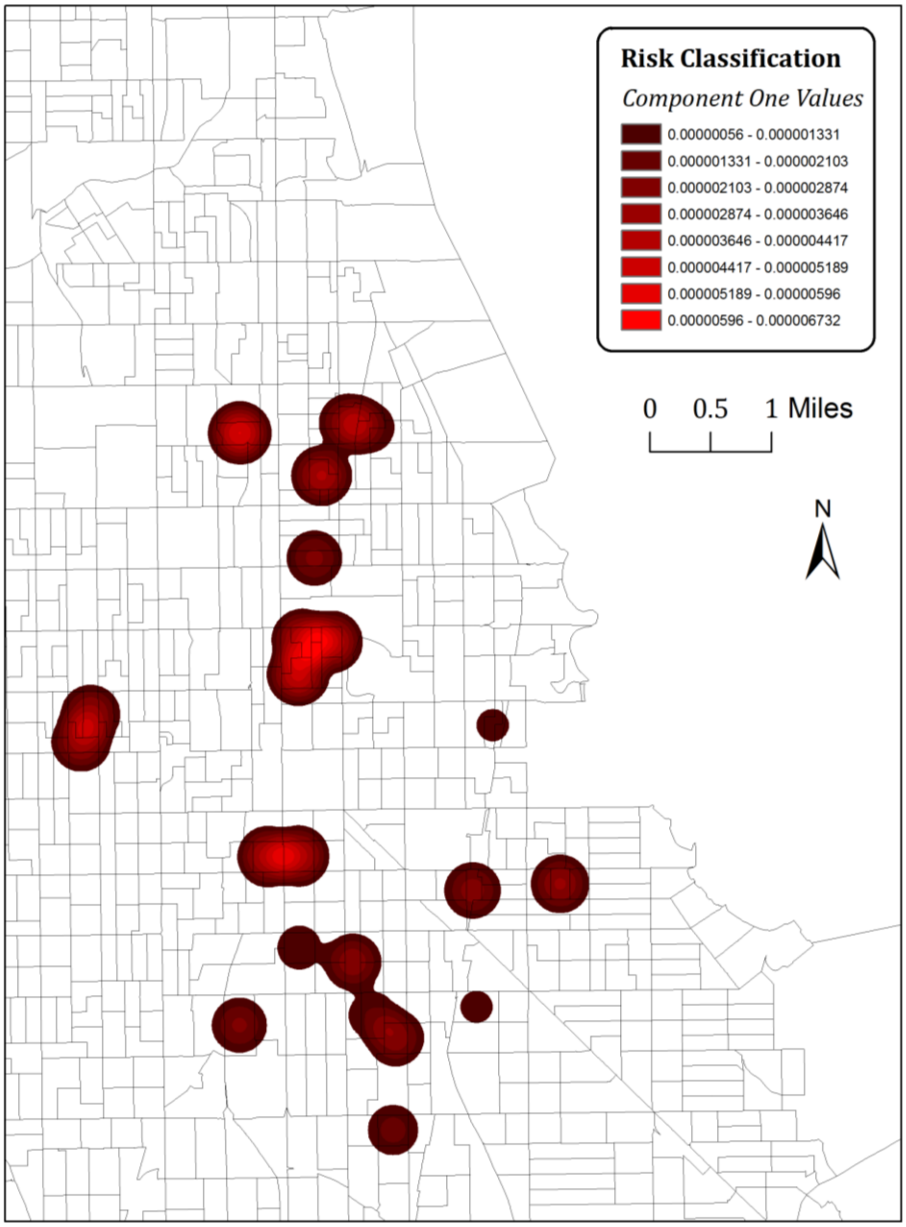

Figure 11. Example of kernel density function that smoothes Component One values between centroids of the 31 Census block groups classified as vulnerable at zero Z-score thresholds.

REFERENCES

Abdi, H. (2003). *Factor rotations in factor analyses*. Program in Cognition and Neurosciences. The University of Texas at Dallas. Dallas.

Abdi, H., & Williams, L. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(4), 433-459. doi: 10.1002/wics.101

Baijal, R., Arora, M., & Ghosh, S. (2006). A GIS assisted knowledge-based approach for military operations (D. o. C. Engineering, Trans.) (pp. 14). Roorkee, India: Indian Institute of Technology.

Baltsavias, E., & Hahn, M. (1999). Integration of Image Analysis and GIS. *International Archives of Photogrammetry and Remote Sensing, 32*, 8.

Cliff, A., & Haggett, P. (1988). Atlas of disease distributions : analytic approaches to epidemiological data [1 atlas (xv, 300 p)].

Cohen, P., & Jensen, D. (1996). *Overfitting Explained*. Paper presented at the Sixth International Workshop on Artificial Intelligence and Statistics.

Cutter, S., Boruff, B., & Shirley, L. (2003). Social Vulnerability to Environmental Hazards. *Social Science Quarterly, 84*(2), 242-261. doi: 10.1111/1540-6237.8402002

de'Donato, F., Stafoggia, M., Rognoni, M., Poncino, S., Caranci, N., Bisanti, L., . . . Perucci, C. (2008). Airport and city-centre temperatures in the evaluation of the association between heat and mortality. *International Journal of Biometeorology, 52*(4), 301-310. doi: 10.1007/s00484-007-0124-5

Dousset, B., Gourmelon, F., Laaidi, K., Zeghnoun, A., Giraudet, E., Bretin, P., . . . Vandentorren, S. (2011). Satellite monitoring of summer heat waves in the Paris metropolitan area. *International Journal of Climatology, 31*(2), 313-323. doi: 10.1002/joc.2222

Dousset, B., Gourmelon, F., & Mauri, E. (2007, 11-13 April 2007). *Application of satellite Remote Sensing for Urban Risk Analysis: a case study of the 2003 extreme heat wave in Paris.* Paper presented at the Urban Remote Sensing Joint Event, 2007.

Harlan, S., Brazel, A., Prashad, L., Stefanov, W., & Larsen, L. (2006). Neighborhood microclimates and vulnerability to heat stress. *Social Science & Medicine, 63*(11), 2847-2863. doi: 10.1016/j.socscimed.2006.07.030

IPCC. (2007). Climate Change 2007: The Physical Science Basis - Summary for Policy Makers. *New York:Cambridge University Press*, 1-18.

Jensen, J. R. (2005). *Remote Sensing of the Environment: An Earth Resource Perspective* (2 ed.): Prentice Hall.

Johnson, D., Wilson, J., & Luber, G. (2009). Socioeconomic indicators of heat-related health risk supplemented with remotely sensed data. *International Journal of Health Geographics, 8*(1), 57.

Kaiser, H. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151. doi: 10.1177/001316446002000116

King-Sun, F., & Rosenfeld, A. (1976). Pattern recognition and image processing. *IEEE Transactions on Computers, C-25*(12), 11.

Klinenberg, E. (2002). *Heat wave : a social autopsy of disaster in Chicago*. Chicago: University of Chicago Press.

Li, C., Goldgof, D., & Hall, L. (1993). Knowledge-based classification and tissue labeling of MR images of human brain. *IEEE Transactions on Medical Imaging, 12*(4), 11.

Lowry, W. (1967). "The Climate of Cities". *Scientific American, 217*(2), 15-23.

Morrow, B. (1999). Identifying and Mapping Community Vulnerability. *Disasters, 23*(1), 1-18. doi: 10.1111/1467-7717.00102

NOAA. (2010). Weather Fatalities Retrieved April 27, 2011, from http://www.weather.gov/om/hazstats.shtml

Oke, T. (1995). *The heat island of the urban boundary layer: characteristics, causes and effects*. Paper presented at the NATO Advanced Study Institute on Wind Climate in Cities, Waldbronn.

Rygel, L., O'sullivan, D., & Yarnal, B. (2006). A Method for Constructing a Social Vulnerability Index: An Application to Hurricane Storm Surges in a Developed Country. *Mitigation and Adaptation Strategies for Global Change, 11*(3), 741-764. doi: 10.1007/s11027-006-0265-6

Semenza, J., Rubin, C., Falter, K., Selanikio, J., Flanders, D., Howe, H., & Wilhelm, J. (1996). Heat-Related Deaths during the July 1995 Heat Wave in Chicago. *New England Journal of Medicine, 335*(2), 84-90. doi: doi:10.1056/NEJM199607113350203

Sheridan, S. (2006). A survey of public perception and response to heat warnings across four North American cities: an evaluation of municipal effectiveness. *International Journal of Biometeorology, 52*, 13.

Stanforth, A. (2011). *Identifying Variations of Socio-spatial Vulnerability to Heat-related Mortality During the 1995 Extreme Heat Event in Chicago, IL, USA.* M.S., IUPUI, Indianapolis.

Voogt, J., & Oke, T. (2003). Thermal remote sensing of urban climates. *Remote Sensing of Environment, 86*(3), 370-384. doi: 10.1016/s0034-4257(03)00079-8

Whitman, S., Good, G., Donoghue, E. R., Benbow, N., Shou, W., & Mou, S. (1997). Mortality in Chicago attributed to the July 1995 heat wave. *Am J Public Health, 87*(9), 1515-1518. doi: 10.2105/ajph.87.9.1515

CURRICULUM VITAE

Kenneth Rulon Simmons

Education

Master of Science, Indiana University-Purdue University Indianapolis
Master of Business Administration, Claremont Graduate University
Bachelor of Science, Brigham Young University

Honors, Awards, and Fellowships

2010 Graduate Fellowship Recipient, U.S. Geospatial Intelligence Foundation
2010 Graduate Fellowship Recipient, Indiana Space Grant Consortium / NASA
2009 Workforce Acceleration Grant Recipient, State of Indiana

Research and Training Experience

Graduate Research Assistant, Department of Geography, Indiana University, Indianapolis, IN
Geospatial Intelligence Internship, National Geospatial-Intelligence Agency, Washington, D.C.
Geographic Information Systems Analysis Internship, The Polis Center, Indianapolis, IN
Geographic Information Systems Analysis Internship, Indianapolis-Marion County Emergency Management Agency, Indianapolis, IN

Professional Experience

NATIONAL GEOSPATIAL-INTELLIGENCE AGENCY, Springfield, VA (June 2010 – Present)
<u>Geospatial Intelligence Analyst</u> (June 2010 – Present)

SIMON PROPERTY GROUP, Indianapolis, IN; Arlington, VA (September 2004-March 2008)
<u>Manager of Mall Marketing</u> (December 2005 – March 2008)
<u>Interim Director of Mall Marketing</u> (October 2005 – November 2005)
<u>Assistant Manager of Mall Marketing</u> (September 2004 - September 2005)

SELF-EMPLOYED MARKETING CONSULTANT, Washington, D.C. (October 2002 – September 2004)
<u>U.S. House of Representatives</u> (client)
<u>Deloitte Touche Tohmatsu</u> (client)

BASKIN-ROBBINS INTERNATIONAL, Glendale, CA (November 1999 – October 2002)
<u>Global Marketing Communications Specialist</u> (July 2001 – October 2002)
<u>International Marketing Coordinator</u> (November 1999 – July 2001)